

2590  
1109

# 2

OIPE

## RAW SEQUENCE LISTING

DATE: 11/06/2001

PATENT APPLICATION: US/09/976,673

TIME: 11:42:32

Input Set : A:\Backup of SEQUENCE LISTING.wbk

Output Set: N:\CRF3\11062001\I976673.raw

ENTERED

```

4 <110> APPLICANT: Lukyanov, Sergey
5   Fradkov, Arcady
6   Labas, Yulii
7   Matz, Mikhail
8   Lukyanov, Konstantin
9   Gurskaya, Nadezda
11 <120> TITLE OF INVENTION: FAR RED SHIFTED FLUORESCENT PROTEINS
14 <130> FILE REFERENCE: CLON-028WO
C--> 16 <140> CURRENT APPLICATION NUMBER: US/09/976,673
17 <141> CURRENT FILING DATE: 2001-10-12
19 <150> PRIOR APPLICATION NUMBER: 60/240,018
20 <151> PRIOR FILING DATE: 2000-10-12
22 <150> PRIOR APPLICATION NUMBER: 60,306,131
23 <151> PRIOR FILING DATE: 2001-07-16
25 <160> NUMBER OF SEQ ID NOS: 26
27 <170> SOFTWARE: FastSEQ for Windows Version 4.0
29 <210> SEQ ID NO: 1
30 <211> LENGTH: 910
31 <212> TYPE: DNA
32 <213> ORGANISM: heteractis crispa
34 <400> SEQUENCE: 1
35 accatttgct ttggttcctt ggcaaacgaa agtttagaac gaaaactgac ccaaattaca 60
36 tcttcctcct ggatccttac catggctggt ttgttgaaag aaagtatgcg catcaagatg 120
37 tacatggaag gcacggttaa tggccattat ttcaagtgtg aaggagaggg agacggcaac 180
38 ccatttacag gtacgcagag catgaggatt catgtcaccg aaggggctcc attaccattt 240
39 gccttcgaca ttttggcacc gtgttgtaga tacggcagca ggaccttgtt ccaccatacg 300
40 gcagagattc cgcatttctt caagcagtc ttccttgaag gctttacttg ggaaagaacc 360
41 acaacctatg aagatggagg cattcttact gctcatcagg acacaagcct ggagggggaa 420
42 tgccttatat acaagtgtaa agtccttggt accaattttc ctgctgatgg ccccgatgat 480
43 aagaacaaat caggaggatg ggagccatgc actgagggtg tttatccaga gaatgggtgc 540
44 ctgtgtggac gtaatgtgat ggcccttaaa gtcggtgatc gtcgtttgat ctgccatctc 600
45 tatacttctt acaggtccaa gaaagcagtc cgtgccttga caatgccagg atttcatttt 660
46 acagacatcc gccttcagat gccgaggaaa acgaaagacg agtactttga actgtacgaa 720
47 gcactctgtg ctaggtacag tgatcttcct gaaaaagcaa attgattgtt cccagtgcac 780
48 ccagactgct gtcagctttt gggttaaagcc cgaagacaaa aaggacattt gtagtttagt 840
49 ttatatcttc ctttcatttg tgaatcaaca ttgtactctc tgtaaaccct taaaatgctc 900
50 cattaaacct                                     910
52 <210> SEQ ID NO: 2
53 <211> LENGTH: 227
54 <212> TYPE: PRT
55 <213> ORGANISM: heteractis crispa
57 <400> SEQUENCE: 2
58 Met Ala Gly Leu Leu Lys Glu Ser Met Arg Ile Lys Met Tyr Met Glu
59 1          5          10          15
60 Gly Thr Val Asn Gly His Tyr Phe Lys Cys Glu Gly Glu Gly Asp Gly
61          20          25          30
62 Asn Pro Phe Thr Gly Thr Gln Ser Met Arg Ile His Val Thr Glu Gly

```

## RAW SEQUENCE LISTING

DATE: 11/06/2001

PATENT APPLICATION: US/09/976,673

TIME: 11:42:33

Input Set : A:\Backup of SEQUENCE LISTING.wbk

Output Set: N:\CRF3\11062001\I976673.raw

```

63          35          40          45
64 Ala Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Pro Cys Cys Glu Tyr
65          50          55          60
66 Gly Ser Arg Thr Phe Val His His Thr Ala Glu Ile Pro Asp Phe Phe
67 65          70          75          80
68 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Thr Thr Thr Tyr
69          85          90          95
70 Glu Asp Gly Gly Ile Leu Thr Ala His Gln Asp Thr Ser Leu Glu Gly
71          100          105          110
72 Asn Cys Leu Ile Tyr Lys Val Lys Val Leu Gly Thr Asn Phe Pro Ala
73          115          120          125
74 Asp Gly Pro Val Met Lys Asn Lys Ser Gly Gly Trp Glu Pro Cys Thr
75          130          135          140
76 Glu Val Val Tyr Pro Glu Asn Gly Val Leu Cys Gly Arg Asn Val Met
77 145          150          155          160
78 Ala Leu Lys Val Gly Asp Arg Arg Leu Ile Cys His Leu Tyr Thr Ser
79          165          170          175
80 Tyr Arg Ser Lys Lys Ala Val Arg Ala Leu Thr Met Pro Gly Phe His
81          180          185          190
82 Phe Thr Asp Ile Arg Leu Gln Met Pro Arg Lys Thr Lys Asp Glu Tyr
83          195          200          205
84 Phe Glu Leu Tyr Glu Ala Ser Val Ala Arg Tyr Ser Asp Leu Pro Glu
85          210          215          220
86 Lys Ala Asn
87 225
90 <210> SEQ ID NO: 3
91 <211> LENGTH: 908
92 <212> TYPE: DNA
93 <213> ORGANISM: heteractis crispa
95 <400> SEQUENCE: 3
96 accatttgct ttggttcctt ggcaaacgaa agtttagacg aaaactgacc caaattacat 60
97 cctcctgatac ttaccatgg ctggtttggt gaaagaaagt atgcgcatca agatgtacat 120
98 ggaaggcacg gttaatggcc attatttcaa gtgtgaagga gagggagacg gcaaccatt 180
99 tacaggtacg cagagcatga ggattcatgt caccgaaggg gctccattac catttgctt 240
100 cgacattttg gcaccgtgtt gtgagtacgg cagcaggacc tttgtccacc atacggcaga 300
101 gattccccgat ttcttcaagc agtctttccc tgaaggcttt acttgggaaa gaaccacaac 360
102 ctatgaagat ggaggcattc ttactgctca tcaggacaca agcctggagg ggaactgcct 420
103 tatatacaag gtgaaagtcc ttggtaccaa tttcctgct gatggccccg tgatgaagaa 480
104 caaatcagaa ggatgggagc catgcaactga ggtggtttat ccagataatg gtgtcctgtg 540
105 tggacgtaat gtgatggccc ttaaagtcgg tgatcgtcgt ttgatctgcc atctctatac 600
106 ttcttacagg tccaagaaag cagtcctgtc cttgacaatg ccaggatttc attttacaga 660
107 catccgcctt cagatgccga ggaaaacgaa agacgagtac tttgaactgt acgaagcatc 720
108 tgtggctagg tacagtgatc ttcttgaaaa agcaaattga ttgttcccag tgacaccaga 780
109 ctgctgtcag cttttggtta aagcccgaag gacaaaagga catttgtagt tttagtttat 840
110 attttccctt tcattttgtg aatcaacatt gtactctctg taaaccttta aaatgctcca 900
111 ttaaacct 908
113 <210> SEQ ID NO: 4
114 <211> LENGTH: 227
115 <212> TYPE: PRT

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/976,673

DATE: 11/06/2001

TIME: 11:42:33

Input Set : A:\Backup of SEQUENCE LISTING.wbk

Output Set: N:\CRF3\11062001\I976673.raw

```

116 <213> ORGANISM: heteractis crispa
118 <400> SEQUENCE: 4
119 Met Ala Gly Leu Leu Lys Glu Ser Met Arg Ile Lys Met Tyr Met Glu
120 1 5 10 15
121 Gly Thr Val Asn Gly His Tyr Phe Lys Cys Glu Gly Glu Gly Asp Gly
122 20 25 30
123 Asn Pro Phe Thr Gly Thr Gln Ser Met Arg Ile His Val Thr Glu Gly
124 35 40 45
125 Ala Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Pro Cys Cys Glu Tyr
126 50 55 60
127 Gly Ser Arg Thr Phe Val His His Thr Ala Glu Ile Pro Asp Phe Phe
128 65 70 75 80
129 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Thr Thr Thr Tyr
130 85 90 95
131 Glu Asp Gly Gly Ile Leu Thr Ala His Gln Asp Thr Ser Leu Glu Gly
132 100 105 110
133 Asn Cys Leu Ile Tyr Lys Val Lys Val Leu Gly Thr Asn Phe Pro Ala
134 115 120 125
135 Asp Gly Pro Val Met Lys Asn Lys Ser Glu Gly Trp Glu Pro Cys Thr
136 130 135 140
137 Glu Val Val Tyr Pro Asp Asn Gly Val Leu Cys Gly Arg Asn Val Met
138 145 150 155 160
139 Ala Leu Lys Val Gly Asp Arg Arg Leu Ile Cys His Leu Tyr Thr Ser
140 165 170 175
141 Tyr Arg Ser Lys Lys Ala Val Arg Ala Leu Thr Met Pro Gly Phe His
142 180 185 190
143 Phe Thr Asp Ile Arg Leu Gln Met Pro Arg Lys Thr Lys Asp Glu Tyr
144 195 200 205
145 Phe Glu Leu Tyr Glu Ala Ser Val Ala Arg Tyr Ser Asp Leu Pro Glu
146 210 215 220
147 Lys Ala Asn
148 225
151 <210> SEQ ID NO: 5
152 <211> LENGTH: 684
153 <212> TYPE: DNA
154 <213> ORGANISM: heteractis crispa
156 <400> SEQUENCE: 5
157 atggctgggtt tgttgaaaga aagtatgcgc atcaagatgt acatggaagg cacggttaat 60
158 ggccattatt tcaagtgtga aggagaggga gacggcaacc catttacagg tacgcagagc 120
159 atgaggattc atgtcaccga aggggctcca ttaccatttg ctttcgacat tttggcaccg 180
160 tgttgtgagt acggcagcag gacctttgtc caccatacgg cagagattcc cgatttcttc 240
161 aagcagtctt tccctgaagg ctttacttgg gaaagaacca caacctatga agatggaggc 300
162 attcttactg ctcatcagga cacaagcctg gaggggaact gccttatata caagtgaaa 360
163 gtccttggtt ccaattttcc tgctgatggc cccgtgatga agaacaaatc aggaggatgg 420
164 gagccaagca ctgaggtggt ttatccagag aatggtgtcc tgtgtggacg taatgtgatg 480
165 gcccttaaag tcggtgatcg tcgtttgatc tgccatctct atacttctta caggtccaag 540
166 aaagcagtcg gtgccttgac aatgccagga tttcatttta cagacatccg ccttcagatg 600
167 ccgaggaaaa cgaaagacga gtactttgaa ctgtacgaag catctgtggc taggtacagt 660
168 gatcttcctg aaaaagcaaa ttga 684

```

## RAW SEQUENCE LISTING

DATE: 11/06/2001

PATENT APPLICATION: US/09/976,673

TIME: 11:42:33

Input Set : A:\Backup of SEQUENCE LISTING.wbk

Output Set: N:\CRF3\11062001\I976673.raw

```

170 <210> SEQ ID NO: 6
171 <211> LENGTH: 227
172 <212> TYPE: PRT
173 <213> ORGANISM: heteractis crispa
175 <400> SEQUENCE: 6
176 Met Ala Gly Leu Lys Glu Ser Met Arg Ile Lys Met Tyr Met Glu
177 1 5 10 15
178 Gly Thr Val Asn Gly His Tyr Phe Lys Cys Glu Gly Glu Gly Asp Gly
179 20 25 30
180 Asn Pro Phe Thr Gly Thr Gln Ser Met Arg Ile His Val Thr Glu Gly
181 35 40 45
182 Ala Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Pro Cys Cys Glu Tyr
183 50 55 60
184 Gly Ser Arg Thr Phe Val His His Thr Ala Glu Ile Pro Asp Phe Phe
185 65 70 75 80
186 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Thr Thr Thr Tyr
187 85 90 95
188 Glu Asp Gly Gly Ile Leu Thr Ala His Gln Asp Thr Ser Leu Glu Gly
189 100 105 110
190 Asn Cys Leu Ile Tyr Lys Val Lys Val Leu Gly Thr Asn Phe Pro Ala
191 115 120 125
192 Asp Gly Pro Val Met Lys Asn Lys Ser Gly Gly Trp Glu Pro Ser Thr
193 130 135 140
194 Glu Val Val Tyr Pro Glu Asn Gly Val Leu Cys Gly Arg Asn Val Met
195 145 150 155 160
196 Ala Leu Lys Val Gly Asp Arg Arg Leu Ile Cys His Leu Tyr Thr Ser
197 165 170 175
198 Tyr Arg Ser Lys Lys Ala Val Arg Ala Leu Thr Met Pro Gly Phe His
199 180 185 190
200 Phe Thr Asp Ile Arg Leu Gln Met Pro Arg Lys Thr Lys Asp Glu Tyr
201 195 200 205
202 Phe Glu Leu Tyr Glu Ala Ser Val Ala Arg Tyr Ser Asp Leu Pro Glu
203 210 215 220
204 Lys Ala Asn
205 225
208 <210> SEQ ID NO: 7
209 <211> LENGTH: 681
210 <212> TYPE: DNA
211 <213> ORGANISM: heteractis crispa
213 <400> SEQUENCE: 7
214 tctggtttgt tgaaagaaag tatgcgcatc aagatgtaca tggaaggcac ggttaatggc 60
215 cattatttca agtgtgaagg agagggagac ggcaacccat ttgcaggtag gcagagcatg 120
216 aggattcatg tcaccgaagg ggtccatta ccatttgcct tcgacatttt ggcaccgtgt 180
217 tgtgagtacg gcagcaggac ctttgtccac catacggcag agattcccga tttcttcaag 240
218 cagtctttcc ctgaaggctt tacttgggaa agaaccacaa cctatgaaga tggaggcatt 300
219 cttactgctc atcaggacac aagcctggag gggaaactgcc ttatatacaa ggtgaaagtc 360
220 cttggtacca attttcctgc tgatggcccc gtgatgaaga acaaatcagg aggatgggag 420
221 ccaagcactg aggtggttta tccagagaat ggtgtcctgt gtggacgtaa tgtgatggcc 480
222 cttaaagtcg gtgatcgctg tttgatctgc catcactata cttcttacag gtccaagaaa 540

```

## RAW SEQUENCE LISTING

DATE: 11/06/2001

PATENT APPLICATION: US/09/976,673

TIME: 11:42:33

Input Set : A:\Backup of SEQUENCE LISTING.wbk

Output Set: N:\CRF3\11062001\I976673.raw

```

223 gcagtcctgt ccttgacaat gccaggattt cattttacag acatccgcct tcagatgctg 600
224 aggaaagaga aagacgagta ctttgaactg tacgaagcat ctgtggctag gtacagtgat 660
225 cttcctgaaa aagcaaattg a 681
227 <210> SEQ ID NO: 8
228 <211> LENGTH: 226.
229 <212> TYPE: PRT
230 <213> ORGANISM: heteractis crispa
232 <400> SEQUENCE: 8
233 Ser Gly Leu Leu Lys Glu Ser Met Arg Ile Lys Met Tyr Met Glu Gly
234 1 5 10 15
235 Thr Val Asn Gly His Tyr Phe Lys Cys Glu Gly Glu Gly Asp Gly Asn
236 20 25 30
237 Pro Phe Ala Gly Thr Gln Ser Met Arg Ile His Val Thr Glu Gly Ala
238 35 40 45
239 Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Pro Cys Cys Glu Tyr Gly
240 50 55 60
241 Ser Arg Thr Phe Val His His Thr Ala Glu Ile Pro Asp Phe Phe Lys
242 65 70 75 80
243 Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Thr Thr Thr Tyr Glu
244 85 90 95
245 Asp Gly Gly Ile Leu Thr Ala His Gln Asp Thr Ser Leu Glu Gly Asn
246 100 105 110
247 Cys Leu Ile Tyr Lys Val Lys Val Leu Gly Thr Asn Phe Pro Ala Asp
248 115 120 125
249 Gly Pro Val Met Lys Asn Lys Ser Gly Gly Trp Glu Pro Ser Thr Glu
250 130 135 140
251 Val Val Tyr Pro Glu Asn Gly Val Leu Cys Gly Arg Asn Val Met Ala
252 145 150 155 160
253 Leu Lys Val Gly Asp Arg Arg Leu Ile Cys His His Tyr Thr Ser Tyr
254 165 170 175
255 Arg Ser Lys Lys Ala Val Arg Ala Leu Thr Met Pro Gly Phe His Phe
256 180 185 190
257 Thr Asp Ile Arg Leu Gln Met Leu Arg Lys Glu Lys Asp Glu Tyr Phe
258 195 200 205
259 Glu Leu Tyr Glu Ala Ser Val Ala Arg Tyr Ser Asp Leu Pro Glu Lys
260 210 215 220
261 Ala Asn
262 225
265 <210> SEQ ID NO: 9
266 <211> LENGTH: 681
267 <212> TYPE: DNA
268 <213> ORGANISM: heteractis crispa
270 <400> SEQUENCE: 9
271 tctggtttgt tgaaagaaag tatgcgcatac aagatgtaca tggaaggcac ggttaatggc 60
272 cattatttca agtgtgaagg agagggagac ggcaacccat ttgcaggtag gcagagcatg 120
273 aggattcatg tcaccgaagg ggtccatta ccatttgctc tcgacatttt ggcaccgtgt 180
274 tgtgcgtacg gcagcaggac ctttgtccac catacggcag agattcccga tttcttcaag 240
275 cagtctttcc ctgaaggctt tacttgggaa agaaccacaa cctatgaaga tggaggcatt 300
276 cttactgctc atcaggacac aagcctggag gggaactgcc ttatatacaa ggtgaaagtc 360

```

VERIFICATION SUMMARY

DATE: 11/06/2001

PATENT APPLICATION: US/09/976,673

TIME: 11:42:34

Input Set : A:\Backup of SEQUENCE LISTING.wbk

Output Set: N:\CRF3\11062001\I976673.raw

L:16 M:270 C: Current Application Number differs, Replaced Current Application Number